

Achieving Higher Frequencies in Large-Scale Nonlinear Model Predictive Control

Victor M. Zavala and Mihai Anitescu

Abstract—We present new insights into how to achieve higher frequencies in large-scale nonlinear predictive control using *truncated-like* schemes. The basic idea is that, instead of solving the full nonlinear optimization (NLO) problem at each sampling time, we solve a single, truncated quadratic optimization (QO) problem. We present conditions guaranteeing stability of the approximation error for truncated schemes using generalized equation concepts. In addition, we propose a preliminary scheme using an augmented Lagrangian reformulation of the NLO and projected successive overrelaxation to solve the underlying QO. This strategy enables early termination of the QO solution because it can perform linear algebra and active-set identification tasks simultaneously. A simple numerical case study is provided.

I. PROBLEM STATEMENT

Consider a nonlinear predictive control (NMPC) problem of the form

$$\min_{u(\tau)} \int_t^{t+N} \psi(z(\tau), u(\tau)) d\tau \quad (1a)$$

$$\text{s.t. } \dot{z}(\tau) = \phi(z(\tau), u(\tau)), \tau \in [t, t+T] \quad (1b)$$

$$z(\tau) \geq 0, u(\tau) \geq 0, \tau \in [t, t+T] \quad (1c)$$

$$z(t) = \bar{z}(t), \quad (1d)$$

where $t \in \mathbb{R}$ is the time dimension, N is the prediction horizon, $u(\cdot) \in \mathbb{R}^{n_u}$ are the control trajectories, and $z(\cdot) \in \mathbb{R}^{n_z}$ are the *model* state trajectories. The nonlinear mapping $\phi(\cdot, \cdot) : \mathbb{R}^{n_z \times n_u} \rightarrow \mathbb{R}^{n_z}$ is the model, and $\psi(\cdot, \cdot) : \mathbb{R}^{n_z \times n_u} \rightarrow \mathbb{R}$ is the cost. This problem is parametric in the *system* states $\bar{z}(\cdot)$, which evolve with unknown dynamics affected by disturbances of different types. The *system* states are assumed to be known at times t_k spaced out by the sampling time $\Delta t = t_{k+1} - t_k$, $k > 0$. We seek to solve the NMPC problem close to *real time* (e.g., with $\Delta t \rightarrow 0$) in order to reject the disturbances and drive the system to the equilibrium. A fundamental problem is that, if the NMPC problem is not solved at a frequency consistent with the system dynamics, the disturbances will accumulate over time and compromise stability. In many applications of interest, achieving high frequencies is not possible because of the computational complexity of the NMPC problem. In other words, the solution time of the NMPC problem limits the attainable Δt . The NMPC problem is typically solved by casting this as a nonlinear optimization (NLO) problem. Here, we consider a parametric NLO of the form

$$\min f(x, t), \quad \text{s.t. } c(x, t) = 0, \quad x \geq 0, \quad (2)$$

where $x \in \mathbb{R}^n$ are the decision variables that include discretized controls and states and, with some abuse of notation, $t \in \mathbb{R}$ will be used to represent the time-evolving data (e.g., system states). A solution of this problem satisfies the parametric Karush-Kuhn-Tucker (KKT) system:

$$x^T \nabla_x \mathcal{L}(w, t) \geq 0, \quad c(x, t) = 0, \quad x \geq 0. \quad (3)$$

This is a nonlinear complementarity system. The Lagrange function is defined as $\mathcal{L}(w, t) = f(x, t) + \lambda^T c(x, t)$, where $\lambda \in \mathbb{R}^m$ are Lagrange multipliers and $w^T = [x^T, \lambda^T]$ with $w \in \mathbb{R}^{n+m}$. We denote the solution of problem as (2) $w^*(t)$. Under certain regularity conditions, this solution forms a continuous but *nonsmooth* manifold [5]. Nonsmoothness arises as a result of variables hitting and moving away from the bounds at subsequent times (active-set changes).

In an effort to achieve higher frequencies, several studies [2], [7], [9] have proposed to solve a single quadratic optimization (QO) problem per sampling time. The QO derived from (2) has the form

$$\min_{\Delta x \geq -\bar{x}_{t_k}} \nabla_x f(\bar{x}_{t_k}, t_{k+1})^T \Delta x + \frac{1}{2} \Delta x^T H(\bar{w}_{t_k}, t_k) \Delta x \quad (4a)$$

$$\text{s.t. } c(\bar{x}_{t_k}, t_{k+1}) + J(\bar{x}_{t_k}, t_k) \Delta x = 0, \quad (4b)$$

where $H(w, t) = \nabla_{xx} \mathcal{L}(w, t)$ is the Hessian of the Lagrange function and $J(x, t) = \nabla_x c(x, t)$ is the Jacobian of the equality constraints. The basic idea is to linearize the KKT conditions at the current point \bar{w}_{t_k} , perturb the data $t_{k+1} = t_k + \Delta t$, solve the QO to obtain the step $\Delta w_{t_k}^T = [\Delta x_{t_k}^T, \Delta \lambda_{t_k}^T]$, and obtain $\bar{w}_{t_{k+1}} = \bar{w}_{t_k} + \Delta w_{t_k}$. The hope is that \bar{w}_{t_k} provides a good approximation to $w^*(t_k)$ for $k > 0$ and that it converges as $k \rightarrow \infty$. An interesting observation arising in NMPC is that the sampling time can be reduced as we decrease the solution time. In the limit, we can expect that $\|\bar{w}_t - w_t^*\| \rightarrow 0$, $\forall t$ as $\Delta t \rightarrow 0$. The traditional approach of converging the NLO to a high degree of accuracy considerably limits the attainable Δt . Therefore, it seems preferable to obtain a fast and sufficiently accurate solution in order to reject disturbances and to *keep neighboring problems close to each other*. Notice, however, that the minimum achievable sampling time Δt is restricted by the solution time of the QO itself. Consequently, some questions arise: For finite Δt , is it possible to guarantee that the error $\|\bar{w}_{t_k} - w_{t_k}^*\|$ remains stable? Can we guarantee $\|\bar{w}_{t_k} - w_{t_k}^*\| \rightarrow 0$ as $k \rightarrow \infty$? How can we accelerate the QO solution to reduce Δt ?

In this work, we address stability issues using generalized equation concepts (Section III). In addition, we propose a

scheme to terminate the QO solution early and still guarantee stability of the error (Section IV). We illustrate these developments using a simple numerical case study (Section V).

II. GENERALIZED EQUATIONS

The key observation driving our analysis is the fact that the complementarity system (3) can be posed as a parametric *generalized equation* (GE) of the following form: For a given $t \in T \subseteq \mathbb{R}$, find $w \in W \subseteq \mathbb{R}^{n+m}$ such that

$$0 \in F(w, t) + \mathcal{N}_K(w). \quad (5)$$

Here, $F : W \times T \rightarrow Z$ is a continuously differentiable mapping in both arguments with

$$F(w, t) = \begin{bmatrix} \nabla_x \mathcal{L}(w, t) \\ c(x, t) \end{bmatrix}, \quad (6)$$

and $K = \mathbb{R}_+^n \times \mathbb{R}^m \subseteq W$ is a polyhedral convex set, where \mathbb{R}_+^n is the non-negativity orthant. We define the derivative mapping $F_w(w, t) := \nabla_w F(w, t)$ and assume that it is Lipschitz in both arguments with constant L_{F_w} , $\forall w \in W, t \in T$. The multifunction $\mathcal{N}_K : W \rightarrow 2^Z$ is the normal cone operator

$$\mathcal{N}_K(w) = \begin{cases} \{\nu \in W \mid (w - y)^T \nu \geq 0, \forall y \in K\} & \text{if } w \in K \\ \emptyset & \text{if } w \notin K. \end{cases}$$

Our goal is to create a discrete-time scheme \bar{w}_{t_k} providing a fast but stable approximation of the solution of (5) $w_{t_k}^*$, $k > 0$. To achieve this, we will perform a single truncated Newton iteration for the generalized equation per time step.

A. The Nonlinear Equation Case

A good intuition as to why a truncated scheme is sufficient to track the solution manifold can be easily explained by considering the case without inequality constraints or, equivalently, when $K = \mathbb{R}^{n+m}$ and $F(w, t) = 0$. In this case, standard calculus results can be used to establish error bounds. This approach has been followed in [2], [9]. In this section, we perform an informal analysis in order to motivate the results of later sections. In the absence of inequality constraints, the approximate scheme \bar{w}_{t_k} , $k > 0$, can be obtained from the recursive solution of the truncated linear Newton system:

$$r_\epsilon = F(\bar{w}_{t_k}, t_{k+1}) + F_w(\bar{w}_{t_k}, t_k)(w - \bar{w}_{t_k}) \quad (7)$$

where r_ϵ is the solution residual satisfying $\|r_\epsilon\| \leq \kappa_\epsilon > 0$. The solution of this problem is denoted as $w = \bar{w}_{t_{k+1}}$. We assume that the linearization point \bar{w}_{t_k} satisfies $\|\bar{w}_{t_k} - w_{t_k}^*\| \leq \kappa_r$ where $F(w_{t_k}^*, t_k) = 0$. In addition, we assume that the solution manifold is Lipschitz continuous (see Theorem 1) such that $\|w_{t_{k+1}}^* - w_{t_k}^*\| \leq L_w \Delta t$ with $\Delta t = t_{k+1} - t_k$ and $\kappa_r, L_w > 0$. We need to establish conditions leading to stability of the approximation error in the sense that

$$\|\bar{w}_{t_k} - w_{t_k}^*\| \leq \kappa_r \quad \Rightarrow \quad \|\bar{w}_{t_{k+1}} - w_{t_{k+1}}^*\| \leq \kappa_r.$$

From the mean value theorem we have that

$$\begin{aligned} 0 &= F(w_{t_{k+1}}^*, t_{k+1}) \\ &= F(w_{t_k}^*, t_{k+1}) \\ &\quad + \int_0^1 F_w \left(w_{t_k}^* + \chi(w_{t_{k+1}}^* - w_{t_k}^*), t_{k+1} \right) (w_{t_{k+1}}^* - w_{t_k}^*) d\chi \end{aligned} \quad (8)$$

and

$$\begin{aligned} F(w_{t_k}^*, t_{k+1}) &= F(\bar{w}_{t_k}, t_{k+1}) \\ &\quad + \int_0^1 F_w (\bar{w}_{t_k} + \chi(w_{t_k}^* - \bar{w}_{t_k}), t_{k+1}) (w_{t_k}^* - \bar{w}_{t_k}) d\chi. \end{aligned} \quad (9)$$

Plugging (7) in (9), we get

$$\begin{aligned} F(w_{t_k}^*, t_{k+1}) &= r_\epsilon - F_w(\bar{w}_{t_k}, t_k)(\bar{w}_{t_{k+1}} - \bar{w}_{t_k}) \\ &\quad + \int_0^1 F_w (\bar{w}_{t_k} + \chi(w_{t_k}^* - \bar{w}_{t_k}), t_{k+1}) (w_{t_k}^* - \bar{w}_{t_k}) d\chi. \end{aligned} \quad (10)$$

From (10) and (8) and bounding we can obtain

$$\begin{aligned} \|\bar{w}_{t_{k+1}} - w_{t_{k+1}}^*\| &\leq \kappa_\psi \kappa_\epsilon \\ &\quad + \kappa_\psi L_{F_w} L_w \Delta t \left(\kappa_r + \frac{1}{2} L_w \Delta t + \Delta t \right) \\ &\quad + \kappa_\psi L_{F_w} \kappa_r \left(\frac{1}{2} \kappa_r + \Delta t \right), \end{aligned}$$

where $\kappa_\psi = \frac{1}{\|F_w(\bar{w}_{t_k}, t_k)\|}$. For stability we require $\|\bar{w}_{t_{k+1}} - w_{t_{k+1}}^*\| \leq \kappa_r$. This implies

$$\begin{aligned} \kappa_r &\geq \kappa_\psi \kappa_\epsilon + \kappa_\psi L_{F_w} L_w \Delta t \left(\kappa_r + \frac{1}{2} L_w \Delta t + \Delta t \right) \\ &\quad + \kappa_\psi L_{F_w} \kappa_r \left(\frac{1}{2} \kappa_r + \Delta t \right). \end{aligned}$$

Rearranging, we have

$$\begin{aligned} \left(1 - \frac{1}{2} L_{F_w} \kappa_\psi \kappa_r \right) \kappa_r &\geq \kappa_\psi \kappa_\epsilon + \kappa_\psi L_{F_w} (L_w + 1) \Delta t \kappa_r \\ &\quad + L_{F_w} \kappa_\psi L_w \left(\frac{1}{2} L_w + 1 \right) \Delta t^2. \end{aligned}$$

Stability follows if $(1 - \frac{1}{2} L_{F_w} \kappa_\psi \kappa_r) > 0$ and if there exist $\kappa \geq 0$ and Δt satisfying

$$\alpha_1^{NLE} \Delta t \kappa_r \leq \kappa \Delta t^2 \quad (11a)$$

$$\alpha_2^{NLE} \Delta t^2 + \kappa_\psi \kappa_\epsilon \leq \alpha_3^{NLE} \kappa_r, \quad (11b)$$

where α_1^{NLE} , α_2^{NLE} , and α_3^{NLE} are defined in the appendix of [13]. At every time t_k , $\bar{w}_{t_{k+1}}$ is obtained by solving (7). This is an approximation of $w_{t_{k+1}}^*$. The stability conditions guarantee that if $\kappa_r, \kappa_\epsilon = O(\Delta t^2)$, then the approximation error remains $O(\Delta t^2)$ for all $k > 0$. We have thus created an algorithm that tracks the solution manifold of $F(w, t) = 0$ stably by solving (within κ_ϵ) a single *truncated* Newton step per sampling time. This allows us to use *iterative linear algebra* schemes that can be terminated early and thus increase the solution frequency of the NMPC problem.

In practice, one can still apply the above results to handle inequality constraints (bounds) by introducing *smoothing* schemes, as suggested in [9], [3], [12]. However, this approach leads to numerical instability. Note also that, in the presence of bounds, we can no longer invert algebraically the Newton system. In addition, nonsmoothness prevents the direct application of standard calculus results. We resolve these technical difficulties in the following sections.

B. Linearized Generalized Equations

An important consequence of the structure of (5) is that it allows us to analyze the smooth and nonsmooth components independently. We start by defining the linearized generalized equation (LGE) at a given solution $w_{t_0}^*$,

$$r \in F(w_{t_0}^*, t_0) + F_w(w_{t_0}^*, t_0)(w - w_{t_0}^*) + \mathcal{N}_K(w). \quad (12)$$

If $K = \mathbb{R}_+^n$, solving this LGE is equivalent to solving the perturbed linear complementarity problem,

$$w \geq 0, \quad \nu = F(w_{t_0}^*, t_0) + F_w(w_{t_0}^*, t_0)\Delta w - r \geq 0, \quad w^T \nu = 0. \quad (13)$$

If F_w is a symmetric matrix, then conditions (13) are the optimality conditions of the QO problem,

$$\begin{aligned} \min_{\Delta w \geq -w_{t_0}^*} & \\ \frac{1}{2} \Delta w^T F_w(w_{t_0}^*, t_0) \Delta w & + F(w_{t_0}^*, t_0)^T \Delta w - r^T \Delta w. \end{aligned} \quad (14)$$

We can rewrite (5) at any point (w, t) in the neighborhood of $w_{t_0}^*$ in terms of (12) by defining the *residual*,

$$r(w, t) = F(w_{t_0}^*, t_0) + F_w(w_{t_0}^*, t_0)(w - w_{t_0}^*) - F(w, t). \quad (15)$$

This gives, for any point satisfying (5),

$$r(w, t) \in F(w_{t_0}^*, t_0) + F_w(w_{t_0}^*, t_0)(w - w_{t_0}^*) + \mathcal{N}_K(w). \quad (16)$$

This formulation will allow us to bound the distance between $(w_{t_0}^*, t_0)$ and neighboring points (w, t) in terms of $r(w, t)$.

Central to this study is the inverse operator $\psi^{-1} : Z \rightarrow W$ of the perturbed LGE (16), which we define as

$$\begin{aligned} w &\in \psi^{-1}[r] \\ \Leftrightarrow r &\in F(w_{t_0}^*, t_0) + F_w(w_{t_0}^*, t_0)(w - w_{t_0}^*) + \mathcal{N}_K(w). \end{aligned} \quad (17)$$

In other words, the operator is a multifunction from the space of the residual (perturbation) of the LGE to the space of the solution. Some basic properties arising from the definition of the inverse operator are as follows:

$$w_{t_0}^* \in \psi^{-1}[r(w_{t_0}^*, t_0)] = \psi^{-1}[0], \quad w_t^* \in \psi^{-1}[r(w_t^*, t)].$$

Definition 1: (Strong Regularity [11]). The GE (5) is said to be strongly regular at $w_{t_0}^*$ in the sense of Robinson if there exists a neighborhood $V_W \subseteq W$ of $w_{t_0}^*$ and a neighborhood $V_Z \subseteq Z$ of $r(w_{t_0}^*, t_0) = 0$, such that for every $r \in V_Z$, (16) has a unique solution $w = \psi^{-1}[r] \in V_W$, and the inverse mapping $\psi^{-1} : V_Z \rightarrow V_W$ is Lipschitz with constant L_ψ . That is, for any $r_1, r_2 \in V_Z$,

$$\|\psi^{-1}[r_1] - \psi^{-1}[r_2]\| \leq L_\psi \|r_1 - r_2\|.$$

This result is a generalization of the implicit function theorem for nonlinear equations. In other words, strong regularity guarantees the invertibility of the solution mapping. In Theorem 4.1 in [11] and Theorem 6 in [4] it has been established that the so-called strong second-order conditions and the linear independence constraint qualification are sufficient to guarantee strong regularity of the NLO.

III. STABILITY OF APPROXIMATION ERROR

Using this basic set of tools, we now establish results that will allow us to construct algorithms for tracking the solution manifold of (5) approximately.

Theorem 1: (Theorem 2.3 in [11] and Theorem 3.3.4 in [5]) Assume (5) is strongly regular at $w_{t_0}^*$. Then, there exist neighborhoods V_W and V_T and a unique and Lipschitz continuous solution $w_t^* \in V_W$ of the GE (5) that satisfies, for each $t = t_0 + \Delta t \in V_T$,

$$(i) \quad \|w_t^* - w_{t_0}^*\| \leq L_w \Delta t \quad (18)$$

with $L_w > 0$. In addition, consider that \bar{w}_t solves the truncated system

$$\delta_\epsilon \in F(w_{t_0}^*, t) + F_w(w_{t_0}^*, t_0)(\bar{w}_t - w_{t_0}^*) + \mathcal{N}_K(\bar{w}_t), \quad (19)$$

where r_ϵ is the solution residual satisfying $\|r_\epsilon\| \leq \delta_\epsilon > 0$. We have that \bar{w}_t satisfies

$$(ii) \quad \|w_t^* - \bar{w}_t\| \leq L_\psi (\delta_\epsilon + \gamma(\Delta t) \Delta t),$$

with $\gamma(\Delta t) \rightarrow 0$ as $\Delta t \rightarrow 0$; and, if F_w is Lipschitz continuous, then

$$(iii) \quad \|w_t^* - \bar{w}_t\| \leq L_\psi (\delta_\epsilon + \kappa \Delta t^2)$$

with $\kappa > 0$.

Having a reference solution $w_{t_0}^*$, we can compute the approximate solution \bar{w}_t by solving the LCP (13) or the QO (14) with $r = F(w_{t_0}^*, t_0) - F(w_{t_0}^*, t)$. From Theorem 1, we see that this approximation can be expected to be close to the optimal solution w_t^* . In our approximate algorithm, however, we relax the requirement that $w_{t_0}^*$ be available. Instead, we consider a linearization point \bar{w}_{t_0} located in the neighborhood of $w_{t_0}^*$. In addition, we assume that the LCP is not solved exactly. In other words, \bar{w}_t is the solution of the truncated system

$$r_\epsilon \in F(\bar{w}_{t_0}, t) + F_w(\bar{w}_{t_0}, t_0)(w - \bar{w}_{t_0}) + \mathcal{N}_K(w), \quad (20)$$

where $r_\epsilon \in \mathbb{R}^n$ is the solution residual. This system can be posed in form (16) by using the following definition:

$$\begin{aligned} r &= r_\epsilon + F(w_{t_0}^*, t_0) + F_w(w_{t_0}^*, t_0)(w - w_{t_0}^*) \\ &\quad - F(\bar{w}_{t_0}, t) - F_w(\bar{w}_{t_0}, t_0)(w - \bar{w}_{t_0}). \end{aligned} \quad (21)$$

Note that, in this case, the perturbation r is an implicit function of the solution $w = \bar{w}_t$. In addition, we emphasize that (21) is used only as an analytical tool. In practice, however, (20) or its corresponding QO is solved. Our first result is given in the following theorem where we establish conditions guaranteeing that the recursive solution of (21) gives rise to a stable tracking scheme of the solution

manifold w_t^* .

Theorem 2: (Stability of Approximation Error). Assume (5) is strongly regular at $w_{t_0}^*$. Define \bar{w}_t as the solution of the perturbed LGE (20) where \bar{w}_{t_0} is a point in the neighborhood V_W of $w_{t_0}^*$. The associated residual $r(\bar{w}_{t_0}, t_0)$ is assumed to satisfy

$$\|r(\bar{w}_{t_0}, t_0) - r(w_{t_0}^*, t_0)\| \leq \delta_r,$$

with $\delta_r > 0$. Assume there exists $\delta_\epsilon > 0$ such that $\|r_\epsilon\| \leq \delta_\epsilon$. If there exists $\kappa > 0$ and if Δt satisfies

$$\alpha_1^{GE} \Delta t \delta_r \leq \kappa \Delta t^2 \quad (22a)$$

$$(\alpha_2^{GE} + \kappa) \Delta t^2 + \delta_\epsilon \leq \alpha_3^{GE} \delta_r, \quad (22b)$$

with $\alpha_1^{GE}, \alpha_2^{GE}$, and α_3^{GE} defined in the appendix of [13], then the approximation error remains stable:

$$\|\bar{w}_{t_0} - w_{t_0}^*\| \leq L_\psi \delta_r \Rightarrow \|\bar{w}_t - w_t^*\| \leq L_\psi \delta_r.$$

Proof. See [13] \square

Corollary 3: Assume conditions of Theorem 2 hold $\forall t_k \in [t_0, t_f]$. Then,

$$\|\bar{w}_{t_k} - w_{t_k}^*\| \leq L_\psi \delta_r, \quad t_{k+1} = t_k + k \cdot \Delta t, \quad \forall k \leq \frac{t_f - t_0}{\Delta t}.$$

Condition (22a) can be satisfied for $\delta_r = o(\Delta t)$, $O(\Delta t^2)$. Condition (22b) is stricter. If $\delta_r = o(\Delta t)$, this condition states that the solution error should be at least $\delta_\epsilon = o(\Delta t)$. Note that a small L_ψ is beneficial because it relaxes both (22a) and (22b). The proposed scheme is equivalent to time-stepping methods used to solve differential variational inequalities (DVI) [10]. Also, note that the above stability results can be applied directly to the NLO context since optimality conditions of QO (4) formulate an LGE of the form (20).

IV. AUGMENTED LAGRANGIAN STRATEGY

As we have seen, solving a single QO (4) at each time step is sufficient. However, it is crucial to have a fast solution strategy for the QO. Here, we propose to reformulate the NLO using an augmented Lagrangian (AL) function and solve the underlying QO using a projected successive over-relaxation (PSOR) strategy. To derive our strategy, we define the AL function,

$$\mathcal{L}_A(x, \bar{\lambda}, t, \rho) = f(x, t) + \bar{\lambda}^T c(x, t) + \frac{\rho}{2} \|c(x, t)\|^2. \quad (23)$$

A strategy to solve the original NLO (2) consists of computing solutions of the AL subproblem

$$\min_{x \geq 0} \mathcal{L}_A(x, \bar{\lambda}, t, \rho) \quad (24)$$

for a sequence of increasing ρ . Note that the multipliers $\bar{\lambda}$ act as parameters of the AL subproblem. The solution of the subproblem is defined as $x^*(\bar{\lambda}, t)$. The multipliers can be updated externally as

$$\bar{\lambda} \leftarrow \bar{\lambda} + \rho c(x^*(\bar{\lambda}, t), t). \quad (25)$$

We thus define the solution pair $x^*(\bar{\lambda}, t)$, $\Lambda^*(\bar{\lambda}, t) = \bar{\lambda} + \rho c(x^*(\bar{\lambda}, t), t)$. The first-order conditions of (24) can be posed as a GE of the form

$$0 \in \nabla_x \mathcal{L}_A(x, \bar{\lambda}, t) + \mathcal{N}_{\mathbb{R}_+^n}(x), \quad (26)$$

where

$$\nabla_x \mathcal{L}_A(x, \bar{\lambda}, t) = \nabla_x f(x, t) + (\bar{\lambda} + \rho c(x, t))^T \nabla_x c(x, t).$$

The linearized version of (26) defined at the NLO solution $x_{t_0}^*$, $\bar{\lambda} = \lambda_{t_0}^*$ is given by the LGE

$$r \in \nabla_x \mathcal{L}_A(x_{t_0}^*, \lambda_{t_0}^*, t_0) + \nabla_{xx} \mathcal{L}_A(x_{t_0}^*, \lambda_{t_0}^*, t_0)(x - x_{t_0}^*) + \mathcal{N}_{\mathbb{R}_+^n}(x) \quad (27)$$

for $r = 0$. To establish perturbation results for the AL LGE (27) in connection with those of the original NLO (2), we consider the following equivalent formulation of (26), proposed in [1]:

$$0 \in F(w, p(\bar{\lambda}), t) + \mathcal{N}_{\mathbb{R}_+^n \times \mathbb{R}^m}(w), \quad (28)$$

where

$$F(w, p(\bar{\lambda}), t) = \begin{bmatrix} \nabla_x f(x, t) + \Lambda^T \nabla_x c(x, t) \\ c(x, t) + p(\bar{\lambda}) + \frac{1}{\rho}(\lambda_{t_0}^* - \Lambda) \end{bmatrix}, \quad (29)$$

$w^T = [x^T \Lambda^T]$, and $p(\bar{\lambda}) = \frac{1}{\rho}(\bar{\lambda} - \lambda_{t_0}^*)$. For $t = t_0$ and $\bar{\lambda} = \lambda_{t_0}^*$, we have $p(\bar{\lambda}) = 0$, $x^*(p(\bar{\lambda}), t) = x_{t_0}^*$, and $\Lambda^*(p(\bar{\lambda}), t) = \lambda_{t_0}^*$. The solution of GE (28) is denoted as $w^*(p(\bar{\lambda}), t)$. The linearized version of (28) at $w_{t_0}^*$ is

$$r \in F(w_{t_0}^*, 0, t_0) + F_w(w_{t_0}^*, 0, t_0)(w - w_{t_0}^*) + \mathcal{N}_{\mathbb{R}_+^n \times \mathbb{R}^m}(w), \quad (30)$$

where

$$F_w(w_{t_0}^*, 0, t_0) = \begin{bmatrix} \nabla_{xx} \mathcal{L}(w_{t_0}^*, t_0) & \nabla_x c(x_{t_0}^*, t_0) \\ \nabla_x^T c(x_{t_0}^*, t_0) & -\frac{1}{\rho} \mathbb{I}_m \end{bmatrix}. \quad (31)$$

We emphasize that the reformulation (28) is considered only for theoretical purposes. In practice, (26) or its corresponding QO is solved. We now establish the following approximation results in the context of the AL framework.

Lemma 4: Assume (26) is strongly regular at $w_{t_0}^*$. Then, there exist neighborhoods V_W, V_T , and V_p where the solution of the AL subproblem (24) satisfies, for each $t = t_0 + \Delta t \in V_T$, $p(\bar{\lambda}) \in V_p$,

$$(i) \quad \|w^*(\bar{\lambda}, t) - w_{t_0}^*\| \leq \frac{L_w}{\rho} \|\bar{\lambda} - \lambda_{t_0}^*\| + L_w \Delta t. \quad (32)$$

Furthermore, consider the approximate solution $\bar{x}(\bar{\lambda}, t)$ obtained from the perturbed LGE (27) with

$$r = \nabla_x \mathcal{L}_A(x_{t_0}^*, \lambda_{t_0}^*, t_0) - \nabla_x \mathcal{L}_A(x_{t_0}^*, \bar{\lambda}, t), \quad (33)$$

and associated multiplier $\bar{\Lambda}(\bar{\lambda}, t) = \bar{\lambda} + \rho c(\bar{x}(\bar{\lambda}, t), t)$. The pair, denoted by $\bar{w}(\bar{\lambda}, t)$, satisfies

$$(ii) \quad \|\bar{w}(\bar{\lambda}, t) - w^*(\bar{\lambda}, t)\| = O\left(\left(\Delta t + \frac{1}{\rho} \|\bar{\lambda} - \lambda_{t_0}^*\|\right)^2\right). \quad (34)$$

Proof. The result follows from the equivalence between (26) and (28), by recalling that $p(\lambda_{t_0}^*) = 0$, $p(\bar{\lambda}) = \frac{1}{\rho} \|\bar{\lambda} - \lambda_{t_0}^*\|$, and by applying Theorem 1. \square

This result states that the solution of a perturbed AL LGE formed at $w_{t_0}^*$ provides a second-order approximation of the subproblem solution $w^*(\bar{\lambda}, t)$. The impact of the multiplier error can be made arbitrarily small by fixing ρ to a sufficiently large value. Our second main result is given in the following theorem. Here, we establish conditions guaranteeing that the recursive solution of the AL LGE gives rise to a stable tracking scheme of the solution manifold w_t^* .

Theorem 5: (Stability of Approximation Error for Augmented Lagrangian). Assume $w_{t_0}^*$ is a strongly regular solution of (27). Define $\bar{x}(\bar{\lambda}, t)$ as the solution of the LGE,

$$r_\epsilon \in \nabla_x \mathcal{L}_A(\bar{x}_{t_0}, \bar{\lambda}, t) + \nabla_{xx} \mathcal{L}_A(\bar{x}_{t_0}, \bar{\lambda}, t_0)(x - \bar{x}_{t_0}) + \mathcal{N}_{\mathbb{R}_+^n}(x), \quad (35)$$

with an associated multiplier update $\bar{\lambda}(\bar{\lambda}, t) = \bar{\lambda} + \rho c(\bar{x}(\bar{\lambda}, t), t)$. The pair is denoted by $\bar{w}(\bar{\lambda}, t)$. The reference linearization point $\bar{w}_{t_0}^T = [\bar{x}_{t_0}^T, \bar{\lambda}_{t_0}^T]$ with $\bar{\lambda}_{t_0} = \bar{\lambda} + \rho c(\bar{x}_{t_0}, t_0)$ is assumed to exist in the neighborhood V_W of $w_{t_0}^*$. The associated residual $r(\bar{w}_{t_0}, t_0)$ is assumed to satisfy $\|r(\bar{w}_{t_0}, t_0) - r(w_{t_0}^*, t_0)\| \leq \delta_r$ with $\delta_r > 0$. Furthermore, assume there exists $\delta_\epsilon > 0$ such that $\|r_\epsilon\| \leq \delta_\epsilon$. If there exists $\kappa > 0$, Δt and ρ satisfying

$$\alpha_1^{AL} \Delta t \delta_r + \frac{L_w}{\rho} \left(\delta_r + \frac{L_w}{L_\psi} \Delta t \right) \leq \kappa \left(\Delta t + \frac{L_\psi \delta_r}{\rho} \right)^2 \quad (36a)$$

$$\alpha_2^{AL} \left(\Delta t + \frac{L_\psi \delta_r}{\rho} \right)^2 + \delta_\epsilon \leq \alpha_3^{AL} \delta_r, \quad (36b)$$

where $\alpha_1^{AL}, \alpha_2^{AL}, \alpha_3^{AL}$ are defined in the appendix in [13], then the approximation error remains stable:

$$\|\bar{w}_{t_0} - w_{t_0}^*\| \leq L_\psi \delta_r \Rightarrow \|\bar{w}(\bar{\lambda}, t) - w_t^*\| \leq L_\psi \delta_r.$$

Proof. See [13]. \square

The recursive stability result of Corollary 3 also applies in this context. Note that if $\rho \rightarrow \infty$, conditions (36a)-(36b) reduce to (22a)-(22b). Therefore, similar order results to those of Theorem 2 can be expected for sufficiently large ρ . Note also that the initial multiplier error (bounded by δ_r) always appears divided by ρ . This indicates that relatively large initial multiplier errors can be tolerated by increasing ρ . Nevertheless, note that the second term on the left-hand side of (36a) remains $o(\Delta t)$ even if $\delta_r = O(\Delta t^2)$. In other words, this condition is more restrictive than (22a).

To solve the QO associated to the AL LGE (35), we follow a PSOR approach. The QO has the form,

$$\min_{z \geq \alpha} \frac{1}{2} z^T \mathbf{M} z + \mathbf{b}^T z. \quad (37)$$

Any solution of this QO solves the LCP,

$$\mathbf{M} z + \mathbf{b} \geq 0, \quad z - \alpha \geq 0, \quad (z - \alpha)^T (\mathbf{M} z + \mathbf{b}) = 0. \quad (38)$$

Consider the following PSOR algorithm adapted from [6], [8]:

PSOR Algorithm

Given $z^0 \geq \alpha$, compute for $k = 0, 1, \dots, n_{iter}$,

$$z_i^{k+1} = (1 - \omega) z_i^k - \frac{\omega}{\mathbf{M}_{ii}} \left(\sum_{j < i} \mathbf{M}_{ij} z_j^{k+1} + \sum_{j > i} \mathbf{M}_{ij} z_j^k - \mathbf{b}_i \right) \\ z_i^{k+1} = \max(z_i^{k+1}, \alpha_i), \quad i = 1, \dots, n, \quad (39)$$

where $\omega \in (0, 2)$ is the relaxation factor.

Theorem 6: (Theorem 2.1 in [8]). Let \mathbf{M} be symmetric positive semidefinite. Then, each accumulation point of the sequence $\{z^k\}$ generated by (39) converges to a solution of the LCP (13). The rate of convergence is R-linear.

It is known that, for the SOR method for linear systems, in order to reduce the error by a factor of 1/10, SOR with non optimal parameter ω requires $O(n)$ iterations [6]. Here, $n = \dim(z)$. We can now establish our algorithm, which we refer to as AugLag.

AugLag Algorithm

Given $\bar{x}_{t_0}, \bar{\lambda}_{t_0}, \Delta t, \rho, n_{iter}$,

- 1) Evaluate $\nabla_x \mathcal{L}_A(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_{k+1}, \rho)$ and $\nabla_{xx} \mathcal{L}_A(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_k, \rho)$.
- 2) Compute step $\Delta \bar{x}_{t_{k+1}}$ by applying n_{iter} PSOR iterations to (37) with:

$$\mathbf{M} = \nabla_{xx} \mathcal{L}_A(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_k, \rho), \\ \mathbf{b} = \nabla_x \mathcal{L}_A(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_{k+1}, \rho)$$

- 3) Update primal variables $\bar{x}_{t_{k+1}} = \bar{x}_{t_k} + \Delta \bar{x}_{t_{k+1}}$ and multipliers $\bar{\lambda}_{t_{k+1}} = \bar{\lambda}_{t_k} + \rho c(\bar{x}_{t_{k+1}}, t_{k+1})$.
- 4) Set $k \leftarrow k + 1$.

The proposed AugLag strategy is attractive because it performs linear algebra and active-set identification tasks simultaneously, it can exploit warm-start information, it can be parallelized, and it has a favorable computational complexity. Note also that the Hessian matrix of the augmented lagrangian remains positive semidefinite close to the solution manifold.

V. NUMERICAL EXAMPLE

To illustrate the developments, we consider the model predictive control of a nonlinear CSTR. The optimal control formulation is given by

$$\min_{u(\tau)} \int_t^{t+T} (w_T(z_T - z_T^{sp})^2 + w_C(z_C - z_C^{sp})^2 + w_u(u - u^{sp})^2) d\tau \\ \text{s.t. } \frac{dz_C}{d\tau} = \frac{z_C - 1}{\theta} + k_0 \cdot z_C \cdot \exp\left[\frac{-E_a}{z_T}\right], \quad z_C(0) = \tilde{z}_C(t) \\ \frac{dz_T}{d\tau} = \frac{z_T - z_T^f}{\theta} - k_0 \cdot z_C \cdot \exp\left[\frac{-E_a}{z_T}\right] \\ + \alpha \cdot u \cdot (z_T - z_T^{cw}), \quad z_T(0) = \tilde{z}_T(t) \\ z_C^{min} \leq z_C \leq z_C^{max}, \quad z_T^{min} \leq z_T \leq z_T^{max} \\ u^{min} \leq u \leq u^{max}.$$

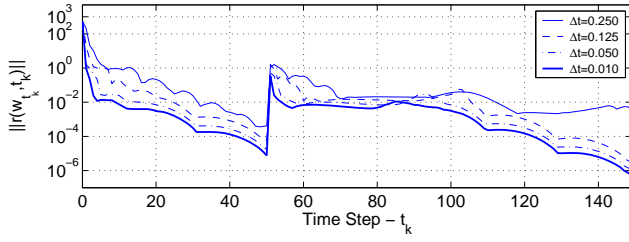


Fig. 1. Residual trajectories for AugLag with increasing Δt .

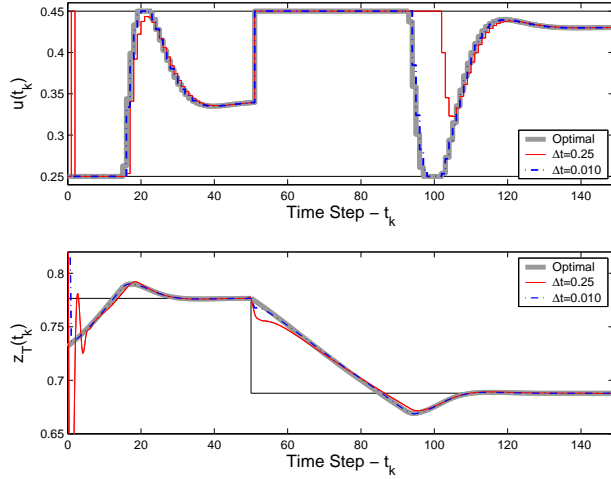


Fig. 2. AugLag and optimal trajectories for the control (top) and temperature (bottom).

The system involves two states $z(\tau) = [z_C(\tau), z_T(\tau)]$ and one control $u(\tau)$. The set-points are denoted by the superscript *sp*. For implementation, the optimal control problem is converted into an NLO by applying an implicit Euler discretization scheme with grid size $\Delta\tau = 0.25$. The NLO is parametric in the initial conditions, which are implicit functions of t . The initial conditions are denoted by $\tilde{z}_T(t)$ and $\tilde{z}_C(t)$. To apply the AugLag algorithm, we set the AL penalty parameter $\rho = 100$ and fix the number of PSOR iterations to 25. We initialize the algorithm by perturbing an initial solution $w_{t_0}^*$ as $\bar{w}_{t_0} \leftarrow w_{t_0}^* \cdot \delta_w$ where $\delta_w > 0$ is a perturbation. This generates the initial residual $r(\bar{w}_{t_0}, t_0)$. An additional perturbation, in the form of a set-point change, is introduced at $t_k = 50$. In Figure 1, we present the norm of the residuals along the simulation horizon with increasing Δt . Note that although the initial residual is large $O(10^3)$, the subsequent residuals remain stable and tend to decrease. The set-point change generates a residual that is only $O(10^0)$ and can be tolerated with no problem. The PSOR residuals r_ϵ at the beginning of the horizon are $O(10^{-1})$ and converge to $O(10^{-6})$ when the system reaches the set-points. In Figure 2, we present control and temperature profiles for $\Delta t = 0.25$ and $\Delta t = 0.01$. The approximation error decreases with the step size. The PSOR strategy identifies efficiently active-set changes in subsequent steps. At a single step, up to 100 changes were observed. For the larger step size, note that

even if the active-sets do not match, the residuals remain bounded and the system eventually converges to the optimal trajectories. We emphasize that these numerical results do not present detailed computational times. However, we can expect that solving the QO problem approximately instead of solving the full NLO problem at each sampling time can reduce the solution frequency by at least an order of magnitude [14], [3].

VI. CONCLUSIONS AND FUTURE WORK

We have presented new insights into enabling the implementation of NMPC at higher frequencies. The main idea is to solve a single, *truncated* quadratic optimization problem per time step. We establish conditions guaranteeing that the approximation error remains stable even in the presence of active-set changes. In addition, we present truncated scheme that enables early termination and that performs linear algebra and active-set identification tasks simultaneously. As part of future work, we are exploring other algorithms with even cheaper steps. In addition, we seek to establish convergence results of approximate schemes and to perform detailed computational studies in large-scale systems.

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy through contract DE-AC02-06CH11357. We thank the reviewers for their questions and suggestions. These were valuable in improving the quality of the manuscript.

REFERENCES

- [1] D. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, 1982.
- [2] M. Diehl, H.G. Bock, and J.P. Schlöder. A real-time iteration scheme for nonlinear optimization in optimal feedback control. *SIAM J. Control and Optim.*, 43:1714–1736, 2005.
- [3] M. Diehl, H. J. Ferreau, and N. Haverbeke. Efficient numerical methods for nonlinear MPC and moving horizon estimation. *In Nonlinear Model Predictive Control*, pages 391–417, 2009.
- [4] A. L. Dontchev and T. R. Rockafellar. Characterizations of strong regularity for variational inequalities over polyhedral convex sets. *SIAM J. Optim.*, 6:1087–1105, 1996.
- [5] J. Guddat, F. Guerra Vazquez, and H. T. Jongen. *Parametric Optimization: singularities, pathfollowing and jumps*. BG Teubner, 1990.
- [6] M. Kočvara and J. Zowe. An iterative two-step algorithm for linear complementarity problems. *Numerische Mathematik*, 68:95–106, 1994.
- [7] W. C. Li and L. T. Biegler. Process control strategies for constrained nonlinear systems. *Ind. Eng. Chem. Res.*, 27:1421–1433, 1988.
- [8] O. L. Mangasarian. Solution of symmetric linear complementarity problems by iterative methods. *JOTA*, 22(4):465–485, 1977.
- [9] T. Ohtsuka. A continuation/GMRES method for fast computation of non-linear receding horizon control. *Automatica*, 40:563–574, 2004.
- [10] J. Pang and D. Stewart. Differential variational inequalities. *Mathematical Programming*, 113:345–424, 2008.
- [11] S. M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research*, 5:43–61, 1980.
- [12] A. G. Wills and W. P. Heath. Barrier function based model predictive control. *Automatica*, 40:1415–1422, 2004.
- [13] V. M. Zavala and M. Anitescu. Real-time nonlinear optimization as a generalized equation. *SIAM J. Control and Optimization*, To Appear, 2010.
- [14] V. M. Zavala, C. D. Laird, and L. T. Biegler. Fast implementations and rigorous models: Can both be accommodated in NMPC? *Int. J. Robust Nonlinear Control*, 18:800–815, 2008.

(To be removed before publication) The submitted manuscript has been created by the University of Chicago as Operator of Argonne National Laboratory ("Argonne") under Contract No. DE-AC02-06CH11357 with the U.S. Department of Energy. The U.S. Government retains for itself, and others acting on its behalf, a paid-up, nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.